GIS Data Layer Design and Creation Guidelines

Last revision: March 5, 2014
Effective date: January 15, 2009

Purpose

This document identifies steps that need to be followed for creation of all Geographic Information Systems (GIS) data sets related to projects of the North Coast and Cascades Network (NCCN) Inventory and Monitoring (I&M) Program. It is envisioned that these guidelines will also be followed in the creation of non-I&M GIS data sets to improve accuracy and usability of all NCCN GIS data.

These guidelines will:

- identify the process by which spatial products are developed and maintained by NCCN GIS staff
- identify Quality Assurance (QA) procedures and data set acceptance criteria to be met when generating and maintaining spatial products

Scope and Applicability

These guidelines should be used by NCCN GIS staff, Project Leads, Data Managers, cooperators, and contractors for creation of spatial data relating to NCCN I&M projects. Specifically, all data that will be maintained on one of the NCCN file servers or submitted to one or more of the national NPS data clearinghouses will be subject to these guidelines. All parties creating and/or submitting GIS-related data to the NCCN should work with NCCN GIS staff in all stages of data creation, along with using guidance within this and other documents listed in the Reference Documents section of this document.

Definitions and Acronyms

Attribute table	A tabular file containing rows and columns. Attribute tables are normally associated with a class of geographic features. Each row represents a geographic feature. Each column represents one attribute of a feature. Feature attribute tables for shapefiles are restricted to dBase (.dbf) format. For geodatabases, attribute tables are contained within the geodatabase folder and cannot be separated from the remainder of the data set other than by an export function.
Cardinality	Term describing how many features in one spatial data set are related to how many records in the attribute table of another data set (i.e., one-to-many, many-to-many, etc.).
Database	A collection of data organized according to a conceptual structure describing the characteristics of the data and the relationships among their corresponding entities. For example, a GIS database includes data about the position and characteristics of spatial features.
Data dictionary	Documentation of a data set that defines attributes such as domains, range limits, and validation rules.
Directionality	Term describing the direction of a relationship between two tables (i.e., source or destination).
Domain	The realm of possible values for a data field. The domain may be constrained either fully or in part by a pick list, range limits and validation rules.
DOQ	Digital Orthophoto Quadrangle

DRG	Digital Raster Graphic
ESRI®	Environmental Systems Research Institute, Inc. of Redlands, CA. Makers of ArcGIS® software, including ArcMap, ArcCatalog and ArcToolbox.
FGDC	Federal Geographic Data Committee. Federal interagency organization that developed the National Spatial Data Infrastructure (NSDI). The NSDI creates policies, standards, and procedures for geographic data production, metadata, and distribution.
Foreign key	A field in a relational table that matches the primary key field of another table. The foreign key can be used to cross-reference tables.
Geodatabase	The ESRI name for ‘geographic database.’ The Geodatabase model is an ArcGIS® version 8.0 and above data format. A geodatabase represents geographic features and attributes as objects and is hosted inside a relational database management system (DBMS).
GIS	Geographic Information Systems. A computer system for creating, storing, checking, integrating, manipulating, analyzing, and displaying spatial data.
GPS	Global Positioning System
I&M	Inventory & Monitoring Program of the National Park Service.
Line	A type of vector geometry having length and direction but no area, connecting at least two x,y coordinate pairs. Lines represent geographic features too narrow to be displayed as an area at a given scale, such as streams, or features with no area that form the boundaries of polygons, such as state and county boundary lines. Lines, or segments of lines, are referred to as “polylines” in ESRI software.
Logical consistency	An evaluation of the interaction between the values of two or more functionally-related attributes. If the values of one attribute change, the values of functionally-related attributes must also change.
Metadata	Data about the data set. Usually provided in the form of a text, xml or html document with information on the data set's quality, coordinate system and projection, attributes, distribution, and citation. In the National Park Service, it is generally a file compliant to the FGDC or ISO Content Standard for Digital Geospatial Metadata and NPS Metadata requirements.
NCCN	North Coast and Cascades Network: http://science.nature.nps.gov/im/units/nccn/
NPS	National Park Service
Physical consistency	An assessment of the topological correctness and geographic extent of the GIS data.
Point	A type of vector geometry used to represent point features. A point is defined by a single x,y coordinate pair.
Polygon	A type of vector geometry used to represent features that have area. A polygon is defined by one or more lines that start and end at the same point.
Orthophoto	An aerial photograph that has been geometrically corrected.
Positional accuracy	A measure of how well each feature’s spatial position in the data set matches reality.
Primary key	A tabular field that uniquely identifies each record in a relational table. It can either be a normal attribute that is guaranteed to be unique or it can be generated by a database management system, such as the globally unique identifier (GUID) created in Microsoft® SQL. Primary keys may consist of a single attribute or multiple attributes in combination.

Random errors	Errors resulting from unknown or accidental causes such as improper recording of spatial or attribute information.
Raster	A spatial data model that defines space as an array of equally sized cells arranged in rows and columns. Raster coordinates are contained in the header of the array. The attribute value of each cell represents the value of the feature the raster represents. Groups of cells that share the same value represent the same type of geographic feature.
Referential integrity	An assessment of the association of related tables based upon their primary and foreign key relationships. Ensures that the primary and foreign keys exist and that all foreign key values match an existing primary key value.
RMS error	Root Mean Square Error. Registration errors are introduced when paper maps, film separates, or digital images are registered to a digitizing board or to images with known coordinate locations. The RMS value represents the amount of error between original and new coordinate locations and is calculated during the digitizing or transformation process.
Shapefile	An ESRI GIS data format that stores non-topological geometry and attribute information for spatial features. The geometry for a feature is stored as a shape comprised of a set of vector coordinates. Shapefiles can support point, line, and area features.
SOP	Standard Operating Procedure.
Systematic errors	Systematic errors are those errors that follow some fixed pattern and are introduced by data collection procedures and systems.
Topology	The spatial relationships between connecting or adjacent spatial features (e.g., polylines, polygons, or points). For example, the topology of a polyline includes its from- and to- vertices and its left and right polygons.
Validity	A means of enforcing attribute accuracy of the database through a set of rules that control data entry.
Vector	A coordinate-based data model that represents geographic features as points, lines, or polygons.

Overview

This document examines steps involved in design, creation, archival, and maintenance of a NCCN I&M data set, as well as the management and planning issues associated with each of the above steps. The quality and usefulness of a final data set depend, almost entirely, on how well the data set was designed. Good data set design allows the data set to be viewed in its entirety so that interaction between elements can be evaluated and permits identification of potential problems and design alternatives. Without a good data set design, there may be data that: will not be used, have no update potential, inappropriately represent entities, lack integration among other parts of the data set, are not supported by other applications, or require major additional costs to revise. Figure 1 summarizes the major steps in data set design and creation, with emphasis given to the data set design process.

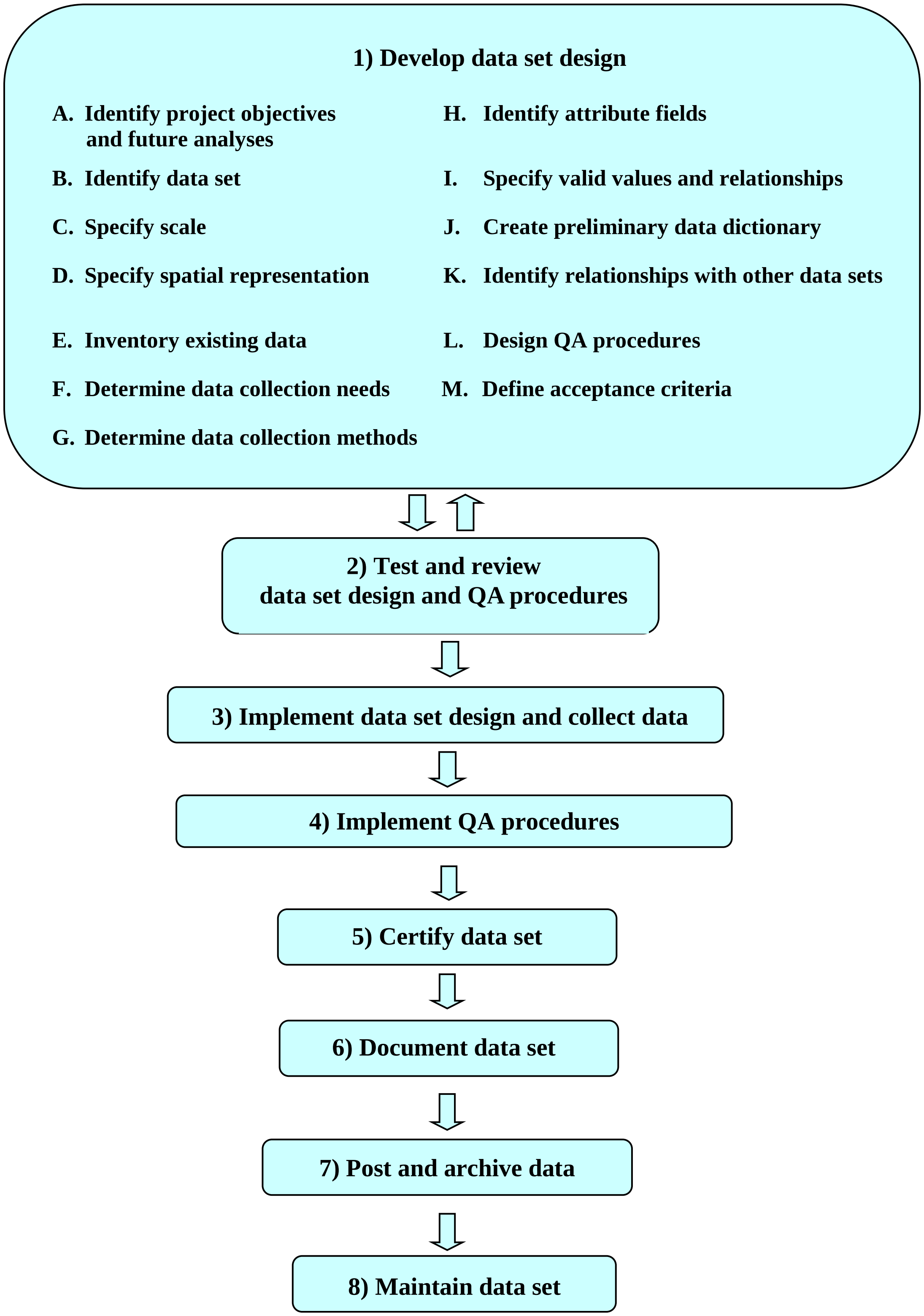

Figure 1. Steps in designing and maintaining an I&M-related spatial data set.

General Procedures

For general guidelines on NCCN project planning and data acquisition, processing, posting, maintenance, and archival, consult the Data Management Plan (Boetsch et al. 2009). More specific information about non-GIS data can also be found in a number of guidance documents posted on the NCCN public website (http://science.nature.nps.gov/im/units/nccn/datamanagement.cfm). The majority of this guidance document has been adapted from documents listed in **Credits** and **Reference Documents** sections at the end of this document.

1. Develop Data Set Design

   A. Identify Project Objectives and Anticipated Analyses
   The Project Lead or Principal Investigator will provide the GIS Specialist with a project proposal stating project hypotheses, objectives, and study design. The Project Lead or Principal Investigator and GIS Specialist will discuss how the new data sets under consideration will be used with existing NCCN data. Appendix 1 describes spatial analysis functions that can be performed on GIS data sets to maximize the utility of spatial information.

   B. Identify Data Set
   Depending on study design, the Project Lead or Principal Investigator, with the help of the GIS Specialist, will determine which spatial data set(s) will be needed for the project. Several data sets might be needed for a single project. For example, an inventory of forest carnivores might need one data set representing potential carnivore habitat, a second one representing a sampling frame (blocks), and a third representing potential locations for camera stations within chosen blocks. In addition, a fourth data set might be created when the coordinates of the actual camera station locations are recorded in the field.

   C. Specify Scale
   The Project Lead or Principal Investigator needs to define the project's spatial accuracy requirements to determine the appropriate scale of data sets to be created. Accuracy requirements should be based on the size of the smallest feature to be mapped and the resolution or scale of other data sets used in conjunction with field data collection. Projects that do not necessitate high accuracy might save money and time by using coarse digitizing methods or by using satellite image interpretation rather than very careful and detailed digitizing or field data collection. National Map Accuracy Standards (US Bureau of the Budget 1947) dictate the accuracy (in feet or meters) on the ground for common scales. Consult GIS Product Specifications (NCCN 2013b) for NCCN I&M project accuracy requirements.

   D. Specify Spatial Representation
   The Project Lead or Principal Investigator, with the help of GIS Specialists, needs to determine whether the data set will be best represented by points, lines, polygons, or rasters. For example, a telephone pole may be represented as a point feature, a road as a line feature, and wetland as an area feature. Spatial representation should be based on types of manipulations that might be undertaken. For example, a building might be represented as a point if building density in a defined area needs to be calculated. A polygon representation is best if one is trying to determine the amount of impervious surface area within a watershed. A raster might be a good alternative to area features if wildlife habitat modeling or watershed analysis will be performed on the data set.

Chosen spatial representation will determine, in part, how the data set is stored and managed. GIS can be used to store and manage both vector and raster data, but vector data (points) can also be stored in a relational database as coordinate pairs. Refer to protocol documents for more guidance on project database design.

E. Inventory Existing Data
Once the needed data sets are identified, the Project Lead or Principal Investigator must coordinate the development of each data set with the GIS Specialist. The GIS Specialist will inventory Park, Network and other agency GIS databases to identify any existing data sets, including relevant Global Positioning System (GPS) data, similar GIS data available at different scales, incomplete GIS data, and other data not currently in a digital format. Appendix 2 provides a list of criteria for which each potential source for a GIS data set should be evaluated. All of the listed criteria need to be documented for each potential data source. If a particular data source is then used to build a GIS data set, some of this information will become part of the permanent metadata.

F. Determine Data Collection Needs
If needed GIS data do not exist or cannot be derived from existing GIS data, the Project Lead or Principal Investigator will determine what spatial information will need to be collected. S/he will work with the GIS Specialist to make sure the collected data will be of appropriate spatial scale and representation (determined in Sections C and D) and compatible with current Park or Network GIS data.

G. Determine Data Collection Methods
There are many means of GIS data collection, which are described in Appendix 3. Gathering field data and data management in a GIS are estimated to be 80% of the cost of a GIS (Trimble 2000). The Project Lead's or Principal Investigator's selection of data collection methods should consider project scale and spatial representation requirements (determined in Sections C and D), Park or Network GIS and GPS capabilities, time requirements for each collection method, and amount of funding available for the data collection and management stages of the project.

H. Identify Attribute Fields
A diverse group of data users should be invited to collaborate on determining the set of attributes collected for each feature. To ensure compatibility among data sets and to meet Park and Network needs, standard attributes should be collected whenever feasible. Consult documentation for specific protocols and GIS Naming Conventions (NCCN 2013a) for NCCN-required core tables and attributes and their domains. Determine for each additional attribute the following:

i. Attribute data type - Are the attributes recorded as numeric, date, time, or character string?

ii. Input restraints - Will the collector need to pick from a list of choices provided in a menu or will it be free text? Should the values be required to be filled in or will the information be considered optional? Should the values be required to fall within a predefined range?

iii. Measurements - What are the units of measurement?

Care should be taken not to use reserved words for attribute names. The list of reserved words is a function of the programming language and will differ with different types and versions of software. Consult the software manufacturer's website or help files to find this information.

If the data set will be merged or combined with data from a different source, make sure the attributes are identical (i.e., the same data type (numeric, character string, etc.) and width) and add an additional attribute to keep track of the source of each feature. In addition, data values should be consistent among data sets (e.g. "Sol Duc River" should be spelled the same way in each instance it appears). Query functions tend to be case-sensitive in GIS software, so values must also have consistent cases.

I. Specify Valid Attribute Values and Relationships

The Project Lead or Principal Investigator, with the help of GIS Specialist and/or Data Manager, is responsible for determining the range or set of values each attribute field within the data set is allowed to have. In addition, s/he is responsible for determining any relationships an attribute field may have to another field.

A value is an assigned or calculated numerical quantity. For each attribute, the Project Lead or Principal Investigator determines the range, pick-list, or type of values the project staff can enter by setting validation rules. Examples of validation rules are: DBH_inch values must be between 0 and 100, Life_stage values must be 'adult' or 'juvenile.'

Another type of validation rule might involve a relationship between two or more fields. Values in an attribute field within a data set often determine values in other attribute fields. For example, if a feature in a landcover classification has a value of "water" in the attribute "Cover_type," then the value of attribute "Slope" is predetermined to be "0" and if it is not recorded as such, represents a logical consistency error.

J. Create a Preliminary Data Dictionary

After inventorying current GIS data sources and determining data collection needs for the project, the Project Lead or Principal Investigator should document the feature and attribute information to be collected in the field or from other existing sources. In this context, "preliminary data dictionary" refers to creating a list of features and their associated definitions and attributes to be collected. Below are the steps to preliminary data dictionary design and implementation. See documentation in a specific protocol and Section 5 of this document for further guidance on creating data dictionaries.

i. Make a list of all the real world physical entities to be mapped.

ii. Categorize the list into point, line, and polygon features.

iii. For each feature, identify the attributes that will be recorded and their data type, input constraints, measurement units, and text description.

iv. For each attribute, identify relationships to other attributes.

At this stage, the GIS Specialist is responsible for creating a GPS data dictionary if GPS data will be collected for the project. The GPS data dictionary should match the design of the project database to facilitate GPS data import into the database. As many attributes as possible should be entered in the field using the GPS data dictionary to reduce data entry time in the office and to minimize the errors that can be introduced when data are transferred from field data sheets into the project database. Field data sheets are still recommended as a backup in case of GPS failure or electronic data loss.

K. Identify Relationships with Other Data Sets (spatial or tabular)
The Project Lead or Principal Investigator, with the help of GIS Specialist and/or Data Manager, must specify how the data sets are related to other data sets, so these relationships can be created in a GIS project or in a relational database.

A relationship is an association between two database objects. These objects can be non-spatial (tables) or spatial (features). For example, a parcel of land (feature) may be associated with an owner (table) or a land-use zone (feature). With a relationship, one can define which field in a feature's attribute table (primary key) and which field in another table (foreign key) share the same values. In this example, the parcel of land contains a Parcel_ID (primary key), which relates to a Parcel_ID (foreign key) in the ownership table and relates to a land-use Zone_ID (foreign key), which relates to the Zone_ID (primary key) in the land-use zone spatial data set. One object can participate in many relationships. Relationships are described based on their directionality and cardinality.

Relationships between spatial and non-spatial data sets can be set up via Open Database Connectivity (ODBC), query layers or table exports. GIS Specialist, Data Manager and Project Lead need to discuss these options as both data sets might require same attribute definitions.

L. Design Quality Assurance (QA) Procedures
The GIS Specialist is responsible for identifying the QA procedures to be followed throughout the GIS data set creation. The selection of QA procedures will depend on the project data collection methods as well as on the desired data accuracy and precision. The following categories of QA should be addressed, if applicable, for each data set and associated tables:

i. Completeness - an indicator of whether each feature or entity is present in the data set, and whether all of its attributes are populated. Completeness as measured by the provider is a relative measure, comparing the data set’s features versus what it is intended to represent. Completeness as assessed by the user is the data set’s fitness for use. See GIS Product Specifications (NCCN 2013b), GIS Naming Conventions (NCCN 2013a) and documentation for a specific protocol for NCCN I&M data set standards.

ii. Attribute Accuracy - describes how well the assigned attribute values match the actual characteristics of the features.

iii. Logical consistency - describes the structural integrity of a data set. This ensures that identified constraints on data keys, attribute domains, and key and attribute interrelationships are met. If the value of one attribute changes, the values of functionally-related attributes might also need to change. For example, in a database in which the attribute SLOPE and the attribute LANDUSE are related, if LANDUSE value is "water," then SLOPE must be 0, as any other value for SLOPE would be illogical.

iv. Physical consistency - the topological correctness and geographic extent of the database. Physical constraints applicable to point data are related to location: for example, do neighboring points violate minimum distance requirements? Tests to verify this constraint can be automated with GIS software, and violations of this type of error may be related to issues of positional accuracy.

Physical constraints applicable to line and polygon data consider what provides a complete and accurate representation of each feature and how the features relate to each other:

1) Are all features completely described graphically?
2) Do any features contain overshoots or undershoots (portions of a line digitized past its intersection with another line or not extended far enough to intersect another line)?
3) Do features intersect only where intended?
4) Do any features exist twice?
5) Are any features too close?
6) Are any polygons too small (sliver)?
7) Do any polygons overlap?
8) Are all features within the intended geographic extent?

v. Referential integrity - the association of related tables based on their primary and foreign key relationships. All tables must have primary and foreign keys and associated sets of data according to predefined rules for each table.

vi. Positional accuracy - measure of how well each spatial object's position in the database matches reality. For example, positional errors can result from incorrect cartographic interpretation, from line segments that were digitized with fewer vertices than necessary, or from inadequate number of significant digits used to store coordinate values. These errors can be random, systematic, and/or cumulative in nature. Positional accuracy must always be documented because the map is a representation of reality.

1) Random errors result from unknown or accidental causes such as improper recording of spatial or attribute information. Random errors will always be a part of any data set, regardless of form. Random error can be reduced by good QA, automated procedures for data entry, and by checking data automatically and visually at various stages in the processing cycle.

2) Systematic errors are those errors that follow some fixed pattern and are introduced by data collection procedures and systems. Systematic errors must be removed by implementation of visual and automated review.

Quality Assurance/ Quality Control Procedures for ITAM GIS Databases (Johnston et al. 1999) describes specific methods for assessment of each of these QA categories.

M. Define Acceptance Criteria

Defining acceptance criteria is probably one of the most difficult parts of a GIS project. It requires knowledge of the data model and database design, as well as the user needs and application requirements, and depends on project schedule, budget, and human resources. Define acceptance criteria for the data set by answering the following questions:

i. Which errors are acceptable?

ii. Are certain errors weighted differently than others?
Each attribute should be reviewed to determine if it is a critical attribute and then weighted accordingly. Primary and foreign keys should not contain any errors, since errors in these attributes will result in errors being propagated throughout the database.

iii. What percentage of error constitutes a rejection of data?
If a GIS data set has 10 features and each feature contains 10 attributes, what is the percentage of error if one attribute of one feature is incorrect? Is it 1 percent or 10 percent? If one subscribes to the theory that all of the attributes must be correct to make a feature correct, then the entire feature is in error, making the error rate 10 percent. On the other hand, if only one attribute is incorrect for a feature, and it is treated as a minor error, then the error rate is 1 percent because one out of a possible 100 attributes is incorrect. Additionally, the positional accuracy of data should be considered. A feature's position, rotation, and scaling must be taken into account when calculating the percentage of error, not just its presence or absence. Consult GIS Product Specifications (NCCN 2013b) for information on NCCN data set accuracy requirements.

iv. Error detection
Once the acceptable percentage of error and the weighting scheme are determined, methods of error detection for data acceptance should be established. These methods are the same as those employed during the data set creation phase (see Sections 1L and 3). If data set was created by digitizing features from an existing base layer, checkplots (a grid of cells overlaying the data set) can be created to compare the features to the original sources. Automated database checking tools can be applied to detect errors in the attribute data. Very large data sets may require random sampling for error detection.

2. Test and Review Data Set Design

The Project Lead or Principal Investigator is responsible for completing a trial run of data collection with the project field crew leads before commencing actual data collection. This will help everyone understand the features and attributes to be collected as well as ensuring familiarity with the equipment and technology. The Project Lead or Principal Investigator can expect to revise the attributes and sometimes the spatial representation of features as the data set design is tested. Below are the steps involved in testing the data set design:

A. Provide necessary training in use of GPS or GIS equipment involved in the data collection stage of the project. GIS Specialist and/or Database Manager must be closely involved in this step.

B. Test the data set design under field conditions, if applicable, by collecting pilot data. Take field crew to a project sample site and collect a set of data, or run through a digitizing or image classification exercise.

C. Check for other features that should be in the data set. Decide if additional useful data can be collected with minimal effort. For instance, if the project was to map an interpretive hiking trail, it will cost little additional effort to map the trail signs as one maps the trail.

D. Evaluate the attribute information collected during the trial run and make sure no necessary attributes were left out.

E. Implement trial QA procedures.

F. Use the pilot data set to create relationships with other data sets or tables outlined in data set design and test their functionality.

G. Make and document any necessary changes to the data set design, data and metadata collection protocols, data entry protocols, QA procedures, or acceptance criteria.

3. Implement Data Set Design and Collect Data
The Project Lead or Principal Investigator is responsible for collecting data necessary for the project. Refer to one of the three sections below, depending on collection method, for instructions on data collection. Regardless of data collection method, field crews are required to QA their data on a systematic basis throughout the field season. For example, the Project Lead or Principal Investigator might designate one day after each week in the field for entering, processing, and/or checking the accuracy of data.

   A. GPS
   Consult GPS Specifications and Guidelines (NCCN 2009), GIS Naming Conventions (NCCN 2013a), and documentation for a specific protocol for GPS data collection specifications.

   B. Tabular Data
   Appendix 4 provides specific instructions for shapefile creation in ArcMap and ArcCatalog from dBase, text and Excel files using Microsoft Excel. Consult documentation for a specific protocol for tabular data standards.

   C. On-screen Digitizing
   Digitize features needed in a new or existing (in some cases) data set. Instructions for creating a shapefile in ArcMap using on-screen digitizing are provided in Appendix 5.

4. Implement QA Procedures

   A. Spatial features
   Implement the following QA procedures (for more specific methods see Johnston (1995)):

      i. Positional accuracy
      Implement visual QA procedures to check for positional accuracy problems. Visually inspecting data can identify systematic errors such as an overall shift in the data caused by an unusually high RMS value or random errors such as missing features. Visual QA can be performed using hard-copy plots or on-screen views. The hard-copy plotting of data is the best method for checking for missing features, misplaced features, and registration errors. On-screen views are an excellent way to verify that edits to the database were made correctly, but are not a substitute for inspecting hard copy plots.

Make sure any erratic or duplicate features have been edited out and alignment problems have been corrected. In cases where high positional accuracy needs to be confirmed or tested, the National Standard for Spatial Data Accuracy (NSSDA) (FGDC 1998) is a good place to start. The NSSDA describes a way to measure and report positional accuracy of features in a spatial data set. Consult GIS Product Specifications (NCCN 2013b) for NCCN I&M data set positional accuracy standards.

ii. Physical consistency
Make sure that all features are topologically correct, i.e. all lines and polygons are properly connected and closed, respectively.

iii. Projection and datum
Upon creation, projection and datum should be specified for all shapefiles (.prj file) and geodatabase feature classes. Add the exported or digitized data sets into ArcMap with other relevant GIS basedata to check for projection and datum problems. All of the data should align properly, within the scale and accuracy constraints of the data sets compared. If the digitized data set does not match basedata, then the collected data were probably assigned the wrong projection or datum. Reexamine the collection and conversion methods to correct any data projection and datum problems.

B. Attribute information
Enter, if necessary, any attribute information collected on field forms or gathered from existing data sources.

i. Validity
Make sure field entries are standardized, complete, and accurate. Visual QA/QC can detect random errors such as typos, presence of erroneous data, or the absence of data. Automated QA/QC procedures can allow quick inspection of large amounts of data. While automated QA/QC procedures may reveal random or systematic errors that were not detected during the visual inspection process, they should be used in addition to, rather than instead of, visual QA/QC procedures. Follow guidelines in the GIS Naming Conventions (NCCN 2013a) document when creating new attributes (fields) to ensure employing the appropriate constraints to prevent many of these errors.

ii. Referential integrity
Make sure primary and foreign keys are present and accurate. No errors should be allowed in these fields, since incorrect information will result in errors cascading through the database and jeopardizing relationships to one or more tables. Perform any advanced data manipulation involving primary and foreign keys to check for errors. This could include combining data with other data sets, creating hotlinks to photos, or performing spatial joins to create a subset of features or add other attribute information.

iii. Logical consistency
Verify that attributes that functionally relate to each other are entered accurately. Correct any errors.

5. Certify Data
After data collection and conversion processes are complete, the Project Lead or Principal

Investigator should review the final data set for data completeness and compliance with accuracy standards and use predetermined criteria for acceptance of the data set as accurate. See documentation for a specific protocol and Project Data Certification Form (NCCN 2006b) for further information.

6. Document Data Set and Data Set Design

   A. Data Dictionary
   After the final data set has been reviewed and accepted, the Project Lead or Principal Investigator is responsible for updating the preliminary data dictionary to reflect any changes that might have occurred during the review process and creating the final data dictionary.

   B. Project Documentation
   Adequate documentation throughout data development is vital for ensuring high quality metadata at the conclusion of a project. The Project Lead or Principal Investigator and GIS Specialist are jointly responsible for documenting their respective parts of the project. There is no standardized project documentation format; however, the Metadata Interview Form is a good starting point (see below).

      i. One way to track project process and progress is to keep a notebook of procedures and output from the project work and/or keep a digital folder to track all project work. The GIS Specialist's and Project Lead's or Principal Investigator's choice on how to track their work depends on the number of participants and complexity of the project. Documentation for a project can range from a simple Readme.txt file to a detailed daily log explaining processes and milestones.
      ii. Documenting project objectives (Section 1A) and creating a preliminary data dictionary (Section 1J) will help the Project Lead or Principal Investigator get started on the project documentation process.
      iii. In addition, the Metadata Interview Form (NCCN 2006a) can help both the GIS Specialist and the Project Lead or Principal Investigator collect the information needed for metadata creation.

   C. Metadata
   The Project Lead or Principal Investigator is responsible for compiling general information about the project and for completing a Metadata Interview Form (NCCN 2006a) for the project. Specific information about data collection methods and data processing should be compiled by the data collectors (field crew leads) and/or data processors (GIS Specialist or Data Manager) for the project. This ensures minimal loss of important information. The Project Lead is responsible for compiling all of the metadata information into an appropriate metadata document. The GIS Specialist (or Data Manager) is then responsible for creating a parsed metadata record from the materials provided by the Project Lead. See How to Create Metadata in ArcGIS 10.0, (GRD 2012) for information on metadata creation and NPS metadata requirements.

7. Post and Archive Data Set
   The GIS Specialist and Data Manager are responsible for posting complete and certified GIS data sets to NPS and NCCN data libraries and repositories. Data archival procedures should be followed as described in the Data Life Cycle section of the Data Management Plan (Boetsch et al. 2009) and in documentation for a specific protocol. Consult the Data Distribution section of the Data Management Plan (Boetsch et al. 2009) for details on the steps required for posting data and the types of data repositories.

8. Maintain Data Set
   The data maintenance stage of the project life cycle begins once the final data set has been accepted by the Project Lead or Principal Investigator and reviewed by the GIS Specialist and the Data Manager. GIS data maintenance involves additions, deletions, and other updates to the data set and metadata. These changes must be done in a tightly-controlled manner in order to retain the database's integrity.

   A. Check-out Procedures
      If a certified and posted data set needs to be edited or updated, the Project Lead or Principal Investigator should contact the GIS Specialist or Data Manager, who will check out the data set from permanent storage, copy it into local storage for updating, and then post it back to the permanent storage location. If more than one person makes changes to the data set and more than one version of the data set exists, the Project Lead or Principal Investigator should work with the GIS Specialist to reconcile the versions before the data set is posted back to permanent storage.

   B. Data Alteration
      Whenever the topology of a feature is altered in a shapefile, attributes such as length and area must be recalculated. Unit values exported from the GPS field data may not match the existing unit values in the GIS layer being edited and should be updated. Note that length and area fields in a geodatabase feature class are automatically updated following feature edits.

   C. Data Merging
      When combining GIS data sets, for example the most recent year’s data being merged with previous year’s, remember that the attribute names in tables need to match exactly in order for the attribute information from all input data sets to be included in the combined data set. The features usually combine properly, but one needs to look carefully at the attribute data to ensure the fields are complete and formatted the same. See GIS Naming Conventions (NCCN 2013a) and documentation for a specific protocol for information on required attributes and standard field formats.

   D. QA Procedures
      Before archiving and posting the data set following updates, QA checks are required to ensure data set integrity. The Data Manager and the GIS Specialist must maintain the data set schema so that table structure and spatial data topologies are not destroyed.

   E. Metadata Update
      The GIS Specialist and Data Manager are responsible for updating the metadata record for the data set following data alteration or data merging. The updated data set is then posted and archived to appropriate NCCN and NPS repositories.

Responsibilities

- Project Leads (NPS), Principal Investigators (contractors and cooperators), and Data Managers working on NCCN I&M projects are responsible for consulting with NCCN GIS Specialists in all stages of GIS data development and, jointly with GIS Specialists, for providing training and support to project staff during I&M GIS data development.
- The Project Lead or Principal Investigator and GIS Specialist will agree on spatial data processing responsibilities before full-scale data collection begins.
- NCCN GIS Specialists and Data Managers are responsible for developing GPS data dictionaries, appropriate QA procedures, and acceptance standards for new GIS data sets.

- The Project Lead or Principal Investigator will verify that all newly-created GIS data sets meet the requirements established in NCCN guidance documents and specific protocols prior to data set submission to the GIS Specialist and/or Data Manager for posting or archiving.

Credits

B. Colonial National Historic Park. 1998. SOP for Themes, Attribute Tables, and Map Design for Colonial NHP, NPS. USDI National Park Service. Available at: http://nps.gov/gis/data_info/documents/colo_gis_sop_95.pdf
C. ESRI. 2013. Data Models. Available at: http://www.esri.com/software/arcgis/geodatabase/data-models
D. ESRI. 2013. The Single-User Geodatabases. Available at: http://www.esri.com/software/arcgis/geodatabase/single-user-geodatabase
E. ESRI. 2013. ArcGIS Help Library. Available at: http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/Welcome_to_the_ArcGIS_Help_Library/00r90000001n000000/
F. NCGIA. 1998. GIS Core Curriculum for Technical Programs. National Center for Geographic Information and Analysis (NCGIA). Available at: http://www.ncgia.ucsb.edu/cctp/
G. Probala, Jeanna. 2005. Surficial Geology Mapping Protocol – Draft. USDI National Park Service, North Coast and Cascades Network.

Recommended Citation

North Coast and Cascades Network – National Park Service. 2013. GIS Data Layer Design and Creation Guidelines. USDI National Park Service.

Contact Information

Natalya Antonova, GIS Specialist
360-854-7312
natalya_antonova@nps.gov

North Cascades National Park Service Complex
810 State Route 20
Sedro-Woolley, WA 98284-1239

Katherine Beirne, GIS Specialist
360-565-3063
katherine_beirne@nps.gov

Olympic National Park
600 East Park Avenue
Port Angeles, WA 98362-6757

Reference Documents

Related Guidance

- Boetsch, J. R., B. Christoe, and R. E. Holmes. 2009. Data management plan for the North Coast and Cascades Network Inventory and Monitoring Program (2005). Natural Resource Report NPS/NCCN/NRR—2009/078. National Park Service, Fort Collins, Colorado. Available at: http://science.nature.nps.gov/im/units/nccn/datamanagement.cfm
- Geographic Resources Division - Intermountain Region – National Park Service. 2012. How to Create Metadata in ArcGIS 10.0. USDI National Park Service. Available at: http://imgis.nps.gov/#Resources.
- Natural Resource Database Template Version 3 Documentation. 2004. Washington Support Office, Inventory and Monitoring Program. National Park Service. Fort Collins, CO. Available at: http://science.nature.nps.gov/im/datamgmt/applications/template/index.cfm.
- North Coast and Cascades Network – National Park Service. 2013a. GIS Naming Conventions. USDI National Park Service. Available at: http://science.nature.nps.gov/im/units/nccn/datamanagement.cfm.

- North Coast and Cascades Network – National Park Service. 2013b. GIS Product Specifications. USDI National Park Service. Available at: http://science.nature.nps.gov/im/units/nccn/datamanagement.cfm.
- North Coast and Cascades Network – National Park Service. 2009. GPS Specifications and Guidelines. USDI National Park Service. Available at: http://science.nature.nps.gov/im/units/nccn/datamanagement.cfm.
- North Coast and Cascades Network – National Park Service. 2006a. Metadata Interview Form. USDI National Park Service. Available at: http://science.nature.nps.gov/im/units/nccn/datamanagement.cfm.
- North Coast and Cascades Network – National Park Service. 2006b. Project Data Certification Form. USDI National Park Service. Available at: http://science.nature.nps.gov/im/units/nccn/datamanagement.cfm.

Other Citations and References

- FGDC. 1998. Geospatial Positioning Accuracy Standards, Part 3: National Standard for Spatial Data Accuracy, FGDC-STD-007.3-1998. Available at: http://www.fgdc.gov/standards/projects/FGDC-standards-projects/accuracy/part3/chapter3.
- Ghose, M. K. Unknown date. Product quality assurance for GIS life-cycle. ISRO, Dept. of Space, Kharagpur. Available at: http://www.gisdevelopment.net/technology/gis/techgi0042.htm.
- Johnston, D. M., D. M. Timlin, D. L. Szafoni and J. J. Casanova. 1999. Quality Assurance/ Quality Control Procedures for ITAM GIS Databases. Geographic Modeling Systems Lab, University of Illinois at Urbana, Champaign. Available at: http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA382487.
- Lillesand, T. M. and R. W. Kiefer. 1994. Remote Sensing and Image Interpretation. John Wiley & Sons, Inc., New York.
- McCain, M. J. and W. C. Masters. 1998. Integrating Quality Assurance into the GIS project life cycle. ESRI ArcUser Magazine. Available at: http://www.esri.com/news/arcuser/798/qa.html.
- NPS GIS Program. 2003. GPS for GIS workflow. USDI National Park Service. Available at: http://www.nps.gov/gis/gps/gps4gis/.
- US Bureau of the Budget. 1947. National Map Accuracy Standards. National Mapping Program Standards, USGS. Available at: http://www.fgdc.gov/standards/projects/FGDC-standards-projects/accuracy/part3/chapter3.

Revision History

Revision Date	Description of Change	Author	Effective Date
Apr. 17, 2007	Updated links to NCCN DM guidance docs	Ronald Holmes	Apr. 17, 2007
January 7, 2009	Updated links to NCCN DM guidance docs, removed ArcView language, removed GIS_Loc_ID references, added information about linking to Access tables via GIS event theme creation	Katherine Beirne	January 15, 2009
December 19, 2012	Updated links and references	Natalya Antonova	

Appendix 1. Spatial analysis with GIS.

- Data modeling (representation)
  Data modeling can involve both vector and raster layers.
    i. Vector representation models describe the features in a landscape such as buildings, streams, or forests using a formalized, conceptual schema, which is usually implemented using points, lines, and polygons. The representation model captures the spatial characteristics of a feature (e.g., the shape of a building) and relationships among features in the landscape (e.g., the distribution of buildings). Along with establishing the spatial relationships, the GIS representation model also models the attributes of the features (e.g., who owns each building). Vector layers can also estimate characteristics of surfaces from a limited number of point measurements. For example, a GIS can quickly generate lines that indicate rainfall amounts, along isoclines, using data from weather stations. The resulting GIS data can be thought of as a rainfall contour layer. This two-dimensional rainfall contour layer may then be overlaid and analyzed with other GIS data covering the same area.
    ii. Raster data are generally divided into two categories, thematic and image. The values in thematic raster data represent some measured quantity or classification of a continuous phenomenon, such as elevation, pollution concentration, or population. For example, in a landcover map, a cell value of 5 may represent forest and a value of 7 may represent water. In contrast, the values of cells in image data represent reflected or emitted light or energy, such as those in a satellite image or scanned photograph.

- Topological modeling
  A GIS can be used to identify and analyze the spatial relationships that exist within digital spatial data. Topological relationships between spatial entities traditionally include adjacency (what adjoins what), containment (what encloses what), and proximity (how close something is to something else). For example, are there any gas stations or factories operating next to the swamp (adjacency)? Are there any within two miles of and uphill from the swamp (proximity)? Such topological relationships allow complex spatial modeling and analysis to be performed.

- Networks
  A GIS can simulate the routing of materials along a linear network. Values such as slope, speed limit, and pipe diameter can be incorporated into network modeling to represent the flow of the materials more accurately. Network modeling is commonly employed in transportation planning, hydrologic modeling, and infrastructure modeling. For example, if all the factories near a wetland were accidentally to release chemicals into the river at the same time, how long would it take for a damaging amount of pollutant to enter the wetland reserve?

- Cartographic modeling (process)
  Cartographic modeling involves using cartographic form (i.e., maps) to depict the interaction of features synthesized in a representation model. It is a technique used for both vector- and raster-based GIS, and it is used to analyze simultaneously both the spatial and thematic characteristics of geospatial information. The thematic component of geospatial information is analyzed via statistical operations on the data (for example, calculating the average and standard deviation of the data), while the spatial characteristics of geospatial information are described through spatial analysis techniques (which are based on coordinate data). Cartographic models are most often used to predict what will happen if some action occurs.
  There are many types of cartographic models including:

i. Suitability modeling: Most spatial models involve finding optimum locations such as finding areas suitable for lynx habitat or areas suitable for logging.
ii. Distance modeling: What is the walking distance between the trailhead and the summit?
iii. Hydrologic modeling: Where will the water flow to?
iv. Surface modeling: What is the geographic extent of the watershed? What is the pollution level for various locations in a watershed?

- Vector overlay
  Vector overlay is the combination of two separate spatial data sets (points, lines, or polygons) to create a new output vector data set. Common vector overlay procedures include Union, Intersection, and Identity operations. All three overlay operations compute the geometric intersection of two vector data sets. Union overlay combines all the geographic features of both input vector data sets. Intersection overlay preserves only those features in the area common to both data sets. Identity overlay preserves all features of the input data set, as well as those features of the identity data set that overlap the input data set. The fields of the attribute tables of the two input vector data sets are combined into the attribute table of the new, output data set.

- Spatial statistics
  Spatial statistics is the collection of statistical methods in which spatial locations play an explicit role in the analysis of data. Spatial data in most cases are not spatially independent. Data values which are close spatially show less variability than data values which are farther away from each other. For example, pollution tends to be concentrated closer to the source and/or natural barriers to dissipation. Spatial statistics can analyze patterns (measure spatial autocorrelation, concentration, and nearest neighbor distance), map clusters, and measure geographic distribution (determine the location of the center of the data, the shape and orientation of the data, and the degree to which features are dispersed). In addition, geostatistics provide means for generating continuous surfaces from selected data points.

- Geocoding
  Geocoding refers to calculating spatial locations (x,y coordinates) from events, which are usually provided in the form of a table or database. A reference theme, such as a stream centerline, is required to geocode individual events or distance ranges. The individual event locations are interpolated, or estimated, by examining distance ranges along a stream segment. For example, in a National Hydrography Dataset stream layer an event point of 500 will be at the midpoint of a specified stream reach that starts with distance 0 and ends with distance 1000. The GIS will then place a point approximately where that event belongs along the segment of centerline.

- Reverse geocoding
  Reverse geocoding refers to estimating an event number from a given set of coordinates. For example, a user can mouse-click on a stream centerline theme (thus providing a set of coordinates), and the GIS will estimate the distance along the stream reach. This distance is interpolated from a range assigned to that stream segment. If the user mouse-clicks at the midpoint of a segment with distances ranging from 1 to 100, the returned value will be approximately 50.

Appendix 2. Evaluation criteria for a data source.

Map series, photos, remotely sensed images, and tabular files are typical sources of data for creating new GIS data sets. Before they can be used as such, they must be reviewed and evaluated for suitability. Critical elements for which these data sources must be evaluated include (note that not all apply to tabular data):

- Appropriate scale
- Projection and coordinate system
- Availability of geodetic control points
- Coverage extent
- Completeness and consistency of feature and attribute data across entire area
- Symbolization of entities (especially positional accuracy of symbols due either to symbol size or off-set placement on map)
- Quality of linework and symbols
- General legibility for digitizing (labels)
- Quality and stability of source material (paper/Mylar)
- Amount of manual editing needed prior to conversion
- Edgematching required between map sheets?
- Existence and type of unique identifiers for each feature - often features shown on map series use so-called "intelligent" keys or identifiers where an identifier for an object contains the map sheet number and/or other imbedded locational codes. For example, a wetland’s unique identifier might contain “H348121” (a USGS map sheet identifier) or “CM” (an abbreviation for USGS Crater Mountain quad). In database design, it is much better to avoid "intelligent" keys of this type.
- Positional and attribute accuracy
- Data source date
- Presence of metadata

Appendix 3. GIS data collection methods

- GPS
  With GPS, positions are collected directly in a digital form in the field, and both spatial features (points, lines, polygons) and attributes are collected (if a data dictionary is loaded on to the GPS unit, thereby eliminating transcription errors). Consult GPS Specifications and Guidelines (NCCN 2009) for NCCN requirements on GPS data collection. Keep in mind that different GPS units collect data with different positional accuracies, and not all GPS units support data dictionaries, which are necessary to record attribute data electronically. These considerations should be addressed when selecting a GPS unit for a project. In addition, field crew leads need to make sure that the GPS unit is properly configured. Project Leads or Principal Investigators (with help from the GIS Specialist) need to make sure that a data dictionary is carefully designed before field data collection begins.

- On-screen digitizing
  On-screen digitizing is an interactive process in which a data set is created using previously digitized or scanned information (such as DOQs) as a background reference. This method of digitizing is commonly called "heads-up" digitizing because the attention of the user is focused up on the computer screen and not on a digitizing tablet. This technique may be used to trace features from a scanned map or image to create new GIS data sets.

- Remote sensing
  Remote sensing is the science and art of obtaining information about an object, area, or phenomenon through the analysis of data acquired by a device that is not in contact with the object, area, or phenomenon under investigation (Lillesand and Kiefer 1994). Remote sensing is accomplished with either passive sensors measuring the reflectance from the earth's surface in certain regions of the electromagnetic spectrum or active sensors measuring radio waves emitted from the sensors themselves. Remote sensing instruments collect raster data that can be further processed to identify objects and classes of interest, such as land cover.

- Photogrammetry
  Soft copy workstations (composed of an image high-resolution scanner; high-speed computer processor; large high-resolution monitor; and appropriate software for viewing scanned images in 3-D, drawing and editing planimetric and topographic information, and translating data sets to various formats for the end user) are used to digitize features directly from stereo pairs of digital photographs. These systems allow data to be captured in 2 and 3 dimensions, with elevations measured directly from a stereo pair using principles of photogrammetry. Currently, analog aerial photos are scanned before being entered into a soft copy system, but as high quality digital cameras are becoming cheaper, this step is often skipped.

- Mobile Devices
  Using mobile devices for spatial data collection is a fast-growing field. Uses have many choices of tablets, smartphones, and embedded devices, such as in-vehicle systems, to collect data using mobile applications that are lightweight, location-aware, and targeted to the work of individual users. Some devices and applications have an ability to synchronize data with NPS servers, often directly from the field. There are restrictions on which mobile devices are available for NPS users. Consult the project Data Manager and/or GIS Specialist for guidance.

Appendix 4. ArcMap: Creating new shapefiles from dBase/text/Excel files using Microsoft Excel spreadsheets.

(Adapted from the NPS Intermountain Geographic Resource Information Management Team, Denver. 2/15/2005. Updated by NCCN on 12/19/2013)

These instructions address how to convert a tabular list of point coordinates into a new ESRI shapefile and save data to a geodatabase feature class. Excel, dBase, and text files can be added directly into ArcMap as event themes. Text files must have a .csv extension to be recognized by ArcMap. Access tables can also be added directly into ArcMap. Access queries cannot be used to create an event theme. Text and dBase files can be opened in Excel and saved as a spreadsheet (.xls) so that further formatting can make data more compatible with ArcMap.

Excel, text, and dBase files

1. Start MS Excel. Open the file with the data and coordinate information by selecting File → Open. Navigate to file location and change file type to "Text Files" or "dBase Files," if necessary. Select file and click "Open." If using a text file, navigate through Text Import Wizard.
   Field headings must at least include: Northing, Easting (for example, Latitude and Longitude or UTM's) and a unique record identification number. See GIS Naming Conventions (NCCN 2013a) for guidelines about field formats and naming conventions. If using geographic coordinates, the format of the coordinates must be in decimal degrees and have a negative sign for west Longitude. For geographic coordinates, it is best to have at least 5 or more significant digits after the decimal to make sure the data are precise enough for large scale mapping.

   A. To convert degrees, minutes, seconds to decimal degrees use the following formula:
      DMS → dd: D + M/60 + S/3600 = dd
      Example: 45°30'45.22"→ 45 + 30/60 + 45.22/3600 = 45.512561

2. Clean up the file. Make hidden columns visible and delete columns that will not be needed in the final GIS shapefile. Remove all empty rows between the header row and data. Empty entries in the coordinate columns will not be used in ArcMap.

3. Make sure the Northing and Easting (or Latitude and Longitude) columns are formatted to 'number'. If there are no decimal places, specify that in the format dialog box (select column Home → Format dropdown menu → Format Cells → Number → Decimal Places). Select all columns, then select Home → Format → Autofit Column Width.

4. Save the file as an Excel file.

Create an Event Theme

5. Open an existing ArcMap document or create a new one. It is important to have the data frame projection defined. Do this by double clicking on the data frame's name (usually 'Layers') in the Table of Contents (TOC, the left pane) portion of ArcMap. Layers → Coordinate System tab

6. Start the dialog box to add XY data by selecting File → Add Data → Add XY Data.

7. The Add XY tool displays the location of XY coordinates from an .xls(x) on-the-fly. Fill in the following sections of the dialog box and then click OK.

A. Name (using the folder button on the right, navigate to the Excel file or Access database and choose the table).

B. Choose the correct X and Y field headings from the table. If the field headings were labeled as Northing/Easting or Latitude/Longitude, the Add XY dialog box automatically fills in the fields for the X and Y coordinates.

C. Set the Spatial Reference (click on the "Edit" button). The program needs to know what projection and datum the coordinates are in.

8. The data will display in the data frame map area as an event layer (not yet a shapefile!). Check to make sure all of the positions are present and that there are no errors in the attributes, if any, or relative locations.

9. Export the data to a shapefile. Right click on the filename in the TOC→ Data → Export Data → Shapefile. Be sure to read the Export Data dialog box carefully for information on the coordinate system to assign to the data. Consult GIS Naming Conventions (NCCN 2013a) for guidelines on naming new GIS files.

Further information can be found in the on-line ArcGIS Resource Center:
http://resources.arcgis.com/en/communities/desktop/
Under Getting started section in the Quick Links on the right side, select an appropriate help page for your version of ArcGIS Desktop.

Appendix 5. ArcGIS: Creating new shapefiles using on-screen digitizing

(Adapted from the NPS Intermountain Geographic Resource Information Management Team, Denver. 9/22/2004. Modified by NCCN for ArcGIS 10.1 on 12/19/2013)

These instructions document how to make a new shapefile data set and populate it by digitizing features on-screen. Both ArcCatalog and ArcMap are used. The instructions can be adapted to create to a geodatabase feature class as well.

1. Open ArcCatalog.

2. Navigate to the location where the new data will be created. This will probably be in a “working” directory on a local machine or in a special folder on a GIS data server.

3. Create a new data set by right clicking in the Contents window (right side panel) or right clicking on the directory name (left side panel) and choosing New → Shapefile. A dialog box appears. Specify the following:

    A. Name of the data set. See GIS Naming Conventions (NCCN 2013a) for guidelines on GIS layer names.
    B. Feature Type (point, polyline, or polygon)
    C. Spatial Reference. Click the “Edit” button, and in the new dialog box, click the “Select” button to find: UTM, zone 10N, meters, NAD83 (NPS and NCCN standard projection).

4. Open ArcMap. Start a new project and save it or open an existing project.

5. Check the data frame’s coordinate system. It needs to be set to the coordinate system that the background and new data set are in. Do this by double clicking on the data frame’s name (usually ‘Layers’) in the TOC portion of ArcMap. Layers → Coordinate System tab.

6. Add any background data needed for guiding relative placement of the new on-screen digitized data. Use the “Add Data” button or drag and drop data from ArcCatalog. Most of the time, this will be raster data such as photos or DRGs. Make sure the background data are in the same coordinate system and datum as the newly created shapefile.

7. Add the new, blank shapefile created in Step 3.

8. Add fields to the attribute table before starting to digitizing features. Right-click on the data set’s name in the TOC → Open Attribute Table to view the attributes. Use the “Option” button to Add Fields. See GIS Naming Conventions (NCCN 2013a) for guidelines on field names and documentation for a specific protocol for information on required fields.

9. Make sure the *Editor* Toolbar is displayed. Do this by going to the Customize → Toolbars → Editor. Once it is open, it can be dragged and “docked” to the side or top of the ArcMap window.

10. Start editing the data by going to *Editor* Toolbar → Editor Menu → Start Editing. Depending on what data are loaded into ArcMap, a dialog box may appear asking which directory holds the data to be edited. Be sure to select the one with the new, blank data set.

11. Zoom to the area of interest where features will be digitized into the new data set.

12. Open the Create Features window by clicking on the Create Features icon on the *Editor* Toolbar (far right). Click on the name of the new shapefile that appears in the top box. If the name of the new shapefile is not there, click on the Organize Templates button → New Template → Check the name of the new shapefile. Finish and close all windows. Select the shapefile in the Create Features window and select the top option in the Construction Tools window. This is the simplest sketch tool.

13. Features can now be traced or drawn in the data frame's map display. Single-click to create points and single-click to start a line or polygon. To end a line or polygon, double-click on the last vertex or right-click and select "Finish Sketch." Note: Save the edits often by going to *Editor* Toolbar → Editor Menu → Save Edits.

14. To view or enter attribute information as features are created, right-click on the data set's name in the TOC → Open Attribute Table. At this point, the newly created feature will be selected (highlighted) in the table. Enter attributes by right-clicking in one of the cells. Alternatively, fill in attributes after digitizing all the features (see #18).

15. To modify a feature, click on the small black arrow tool on the *Editor* Toolbar, and click on the feature to be modified to select it. Then select the Edit Vertices or Reshape Feature Tool on the *Editor* Toolbar.

16. To delete a feature, select it with the small black arrow tool on the *Editor* Toolbar and then hit the keyboard's delete key.

17. When done adding and modifying data, go to the *Editor* Toolbar → Editor Menu → Save Edits and then → Stop Editing.

18. To edit or add data to the shapefile's tabular data, right-click on the data set's name in the TOC → Open Attribute Table. Note: Use the "Table Options" button to Add Fields when the data set is *not* in editing mode. The new fields can be populated once editing is resumed (see Step 10).

For more information on editing, consult the ArcGIS on-line Resource Center at http://resources.arcgis.com/en/communities/desktop/ or ArcGIS Desktop Help files.